

## ***In vivo* precision assessment of a near-infrared spectroscopy-based tissue oximeter (OxyPrem v1.3) in neonates considering systemic hemodynamic fluctuations**

Stefan Kleiser  
Daniel Ostojic  
Nassim Nasser  
Helene Isler  
Hans Ulrich Bucher  
Dirk Bassler  
Martin Wolf  
Felix Scholkmann  
Tanja Karen

# ***In vivo* precision assessment of a near-infrared spectroscopy-based tissue oximeter (OxyPrem v1.3) in neonates considering systemic hemodynamic fluctuations**

**Stefan Kleiser,<sup>a,\*</sup> Daniel Ostojic,<sup>a</sup> Nassim Nasser, <sup>a,b</sup> Helene Isler,<sup>a</sup> Hans Ulrich Bucher,<sup>a</sup> Dirk Bassler,<sup>a</sup> Martin Wolf,<sup>a</sup> Felix Scholkmann,<sup>a,b,†</sup> and Tanja Karen<sup>a,†</sup>**

<sup>a</sup>University of Zurich, University Hospital Zurich, Biomedical Optics Research Laboratory, Department of Neonatology, Zurich, Switzerland

<sup>b</sup>University of Bern, Institute of Complementary Medicine, Bern, Switzerland

**Abstract.** The aim was to determine the precision of a noninvasive near-infrared spectroscopy (NIRS)-based tissue oximeter (OxyPrem v1.3). Using a linear mixed-effects model, we quantified the variability for cerebral tissue oxygenation (StO<sub>2</sub>) measurements in 35 preterm neonates to be 2.64%, a value that meets the often-articulated clinicians' demand for a precise tissue oxygenation measurement. We showed that the variability of StO<sub>2</sub> values measured was dominated by spontaneous systemic hemodynamic fluctuations during the measurement, meaning that precision of the instrument was actually even better. Based on simultaneous and continuous measurements of peripheral arterial oxygenation and cerebral StO<sub>2</sub> with a second sensor, we were able to determine and quantify the physiological instability precisely. We presented different methods and analyses aiming at reducing this systematic physiological error of *in vivo* precision assessments. Using these methods, we estimated the precision of the OxyPrem tissue oximeter to be  $\leq 1.85\%$ . With our study, we deliver relevant information to establish highly precise cerebral oxygenation measurements with NIRS-based oximetry, facilitating the further development toward a substantially improved diagnosis and treatment of patients with respect to brain oxygenation. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.23.6.067003](https://doi.org/10.1117/1.JBO.23.6.067003)]

Keywords: near-infrared spectroscopy; cerebral oximetry; precision; preterm infant.

Paper 180149R received Mar. 12, 2018; accepted for publication May 21, 2018; published online Jun. 14, 2018.

## **1 Introduction**

Infants born preterm are vulnerable to hypoxic and ischemic cerebral insults that can lead to long-term morbidity.<sup>1</sup> Diagnostic methods to early detect these conditions and to prevent lesions are urgently needed. Optical methods such as near-infrared spectroscopy (NIRS) have the potential to fulfill this need by assessing cerebral oxygenation and blood perfusion (hemodynamics) noninvasively, especially by being able to measure the tissue oxygenation (StO<sub>2</sub>) as an absolute parameter.<sup>2,3</sup> The measurement of cerebral StO<sub>2</sub> may be useful for detecting situations where the oxygenation of the brain is impaired; the hope is that by properly adjusting the cerebral oxygenation, brain lesions will be prevented. However, the *in vivo* precision of NIRS-based oximetry has repeatedly been reported to be insufficient for clinical practice.<sup>2,3</sup>

In previous studies, the reproducibility of StO<sub>2</sub> was examined for various NIRS-based oximeters. For example, Sorensen and Greisen<sup>4</sup> reported a precision of 5.2% for the NIRO 300 (Hamamatsu Photonics, Hamamatsu, Japan). Recent studies showed that a higher precision in neonates of 2.0% (term) to 4.2% (preterm) can be achieved by excluding inhomogeneous tissue for the OxiplexTS (ISS, Champaign, Illinois).<sup>5</sup>

Technically, precision is quantified by the variation of repeated measurements with an instrument at an unchanged specimen, i.e., the true value of StO<sub>2</sub> is assumed to be constant. In studies *in vivo*, this is not necessarily given as numerous regulation processes may cause physiological changes and thus changes in true StO<sub>2</sub>. These physiological changes contribute to the observed variability, causing it to be higher than the variation caused by the device (i.e., the precision). However, vital parameters associated with changes in true StO<sub>2</sub>, such as heart rate or arterial oxygen saturation (SaO<sub>2</sub>), were unfortunately not monitored during the measurements in previous studies, such as Refs. 5 and 6. Especially, SaO<sub>2</sub> variability matters because it acts as an “input parameter,” and changes in SaO<sub>2</sub> are directly reflected in StO<sub>2</sub>.<sup>6</sup>

Data on the influence of these physiological changes or spontaneous systemic hemodynamic fluctuations on precision of NIRS measurements are sparse and inconclusive. While Hyttel-Sorensen et al.<sup>7</sup> found that on the adult forearm most of the variation was not due to spontaneous fluctuations, Menke et al.<sup>8</sup> stated that “most of the variation between repeated rSO<sub>2</sub> measurements is due to physiological variation” in their measurements on the forehead of preterm and term neonates.

The aim of our study was first, to determine the precision of the OxyPrem NIRS-based oximeter device in preterm neonates and second, to assess physiological changes during such measurements. Furthermore, we aimed to present several approaches reducing the impact of spontaneous systemic hemodynamic fluctuations on the measurements.

\*Address all correspondence to: Stefan Kleiser, E-mail: [stefan.kleiser@usz.ch](mailto:stefan.kleiser@usz.ch)

†Shared last authorship

**Table 1** Demographic data of preterm infants ( $n = 35$ ).

	Median	Range
Gestational age (weeks)	33(2/7)	25(1/7) to 36(1/7)
Postmenstrual age (weeks)	35(5/7)	32(1/7) to 38(5/7)
Birth weight (g)	1630	730 to 2820
Weight at measurement (g)	2070	1460 to 3000

## 2 Methods

### 2.1 Subjects

The study was performed at the Department of Neonatology, University Hospital Zurich. The study protocol was accepted by the ethical committee of Zurich (KEK 2010-0102/2) and received regulatory approval from Swissmedic (2010-MD-0019). Parental consent was obtained in all cases prior to enrollment.

Thirty-seven clinically stable preterm infants breathing spontaneously on room air were enrolled in our study. The demographic data of the evaluated infants are shown in Table 1. Infants with congenital malformations were excluded.

### 2.2 Instrumentation

StO<sub>2</sub> was measured with our in-house built NIRS-based oximeter OxyPrem v1.3, peripheral arterial oxygen hemoglobin saturation (SpO<sub>2</sub>) was acquired by a Sensmart Model X-100 Universal Oximetry System with “6100CN cloth sensor—neonatal” pulse oximetry sensor (Nonin Medical, Inc., Plymouth, Minnesota).

The OxyPrem v1.3 sensor consists of four continuous-wave light sources and two light detectors [Fig. 1(c)], linearly and symmetrically arranged with sources placed in between the detectors. Each source is equipped with multiple LEDs (690, 760, 805, and 830 nm). The two detectors D1, D2, and sources S1, S4 form an outer region of interest (ROI<sub>1</sub>) with source-detector separations (SDS) of 15 and 35 mm, respectively. Both detectors and sources 2 and 3 form an inner region

(ROI<sub>2</sub>) with SDS of 20 and 30 mm, respectively. Except for this difference in source positions and SDS, both ROIs are technically identical and the major difference is that slightly different tissues are probed with ROI<sub>1</sub> offering deeper tissue sensitivity (due to the longer SDS). Sources and detectors are embedded in flexible silicone, and the sensor's shape is adapted for easy application to strongly curved surfaces such as the tiny head of neonates.

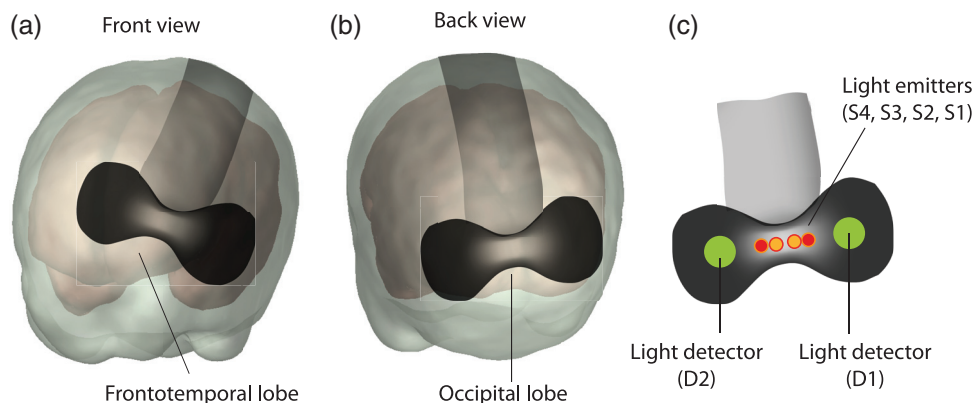
The multidistance algorithm is based on the following steps: the ambient light intensity is subtracted. From the slope of intensity decrease over distance, the effective attenuation is calculated which is then transformed into concentrations of oxyhemoglobin and deoxyhemoglobin. Finally, StO<sub>2</sub> is calculated as the ratio of oxyhemoglobin over the sum of oxyhemoglobin and deoxyhemoglobin. For each ROI, StO<sub>2</sub> is calculated individually. The mean of both StO<sub>2</sub> values for ROI<sub>1+2</sub> StO<sub>2</sub> is calculated as well.

We employ a self-calibrating algorithm,<sup>9</sup> which exploits the sensor's symmetry to cancel most influential factors. For instance, differences in light coupling between sources of detectors are automatically canceled to a high degree. Mainly, this algorithm enables OxyPrem to compensate for local inhomogeneities such as an unequal distribution of hair underneath the sensor, birth marks, hematoma, smears of blood, or meconium under a source or detector. This algorithm also increases robustness to motion artifacts.<sup>10</sup> OxyPrem thus provides a robust StO<sub>2</sub> measurement.

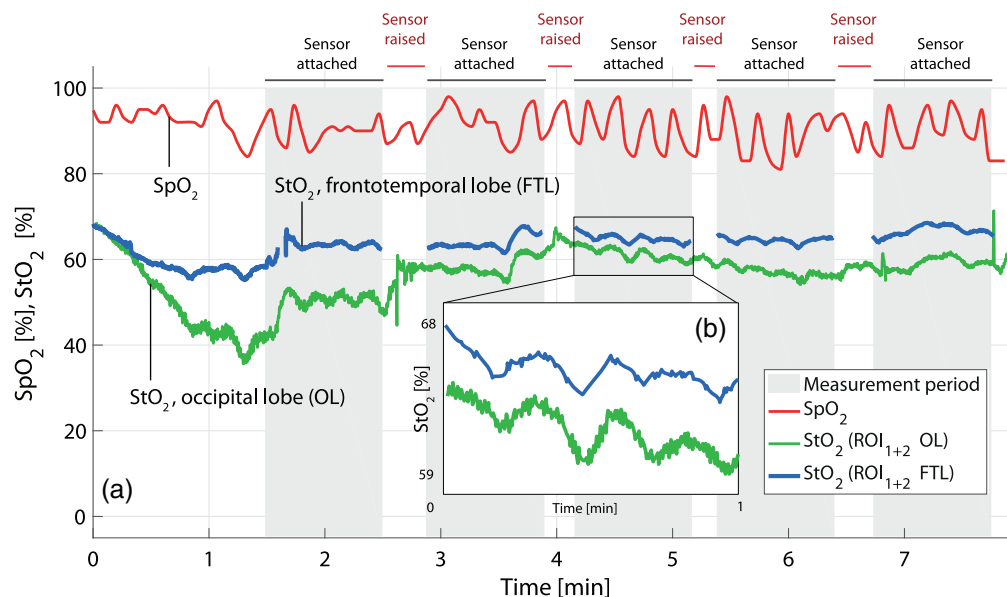
In addition, in contrast to many other oximeters, OxyPrem's StO<sub>2</sub> is insensitive to the total hemoglobin concentration.<sup>11</sup> Although designed for cerebral oximetry in neonates, it has a similar capability to assess the oxygenation in deeper layers of tissue as common adult sensors<sup>12</sup> and is therefore a tool that determines StO<sub>2</sub> in a variety of patients and anatomical locations.

### 2.3 Data Acquisition

Five 1-min measurements were conducted with an OxyPrem sensor placed over the left frontotemporal lobe (FTL) on the head of a preterm neonate. In between, the sensor was removed and reattached (resited) at approximately the same location without marking. The FTL was selected because it is a common region to place NIRS sensors as it is not covered by hair. Before



**Fig. 1** Visualization of the OxyPrem sensor placements (a) FTL, (b) OL and the sensor design in bottom view: (c) light detectors (D) in green and light sources (S) in red (ROI<sub>1</sub>)/orange (ROI<sub>2</sub>). From right to left: D1, S1, S2, S3, S4, and D2. ROI<sub>1</sub> is constituted by D1, D2, S1, and S4. ROI<sub>2</sub> is constituted by D1, D2, S2, and S3. The neonatal head model is based on a fMRI scan of a 35-week-old infant.



**Fig. 2** Typical time series of one patient (ID 12) showing the multiply removed and reattached (resited) FTL and fixed OL StO<sub>2</sub> and SpO<sub>2</sub> signals. The gray backgrounds indicate the 1-min measurement periods.

the start of the next measurement, we waited for sensor LEDs to regulate (see Fig. 2 for a typical example).

To assess the spontaneous systemic hemodynamic fluctuations, we recorded SpO<sub>2</sub> on the right arm and placed a second OxyPrem sensor over the occipital lobe (OL) with the center of the sensor being positioned 1 cm above the inion. The OL was selected as a reference location because it is far away from the FTL and thus avoids cross talk between the two sensors. To obtain data with high signal-to-noise ratio, we recruited preferably light-haired subjects. The placement of the OxyPrem sensors over the head of a neonate is shown in Figs. 1(a) and 1(b).

## 2.4 Data Analysis

The StO<sub>2</sub> and SpO<sub>2</sub> signals were synchronized by setting an event mark recorded directly after having started the recording. Data selection and analysis were performed as follows:

- I. Typically five median values (one for each 1-min measurement, called “block median”) per subject were analyzed by a linear mixed-effects model with subject as factor (LMEM<sup>13,14</sup>) using the statistics software R (Version 3.3.2).<sup>15</sup> The average residual corresponds to the variability within subjects and is therefore called intrasubject standard deviation (SD<sub>intraSubj</sub>). This approach has been used in previous studies, for instance, in Ref. 16. We evaluated StO<sub>2</sub> of ROI<sub>1</sub> and ROI<sub>2</sub> individually as well as their average StO<sub>2</sub> (ROI<sub>1+2</sub>) and included all data, irrespective of artifacts caused by drops in SpO<sub>2</sub> or subject movement (“all available”).
- II. Visual inspection of the StO<sub>2</sub> data revealed several subjects in which  $\geq 3$  out of 5 blocks were containing artifacts, mostly motion artifacts or desaturations. These subjects were excluded for further analysis with the LMEM (“manual rating”). This approach was

inspired by functional studies, where artifacts are often rated manually.

- III. In another analysis with an LMEM, subjects showing high physiological variation between block medians were removed. The criterion for subject removal was an SD between the five block medians of  $> 2\%$  for SpO<sub>2</sub> (“stable SpO<sub>2</sub>”). SpO<sub>2</sub> reflects systemic physiology and hence thresholds for variability in SpO<sub>2</sub> lead to a dataset with smaller physiological changes, but also less data for analysis. Our thresholds were selected as a trade-off between removing physiologically unstable subjects and retaining data. We set a strict SpO<sub>2</sub> threshold because changes in SpO<sub>2</sub> lead to concurrent changes in StO<sub>2</sub> anywhere in the body.<sup>7</sup>
- IV. In this dataset, an additional criterion for stable OL StO<sub>2</sub> was added (SD  $< 7.5\%$ ), to exclude drifts in cerebral StO<sub>2</sub> which are possible even if SpO<sub>2</sub> is stable. The threshold was set to remove only the most severe cases, i.e., it removed only one, severely affected subject.
- V. and VI. Two additional analyses were performed, requiring variability of StO<sub>2</sub> within a block to be below a threshold. Variability during the blocks was determined by taking the median absolute deviation (MAD<sup>17</sup>) of all samples within a block. We chose MAD as a measure because it is more robust against outliers than the SD. For better comparability, a factor of 1.4826 was applied (scaled MAD<sup>18</sup>) to obtain similar values to SD for artifact free, normally distributed data. We then compared scaled MAD to two different thresholds:  $< 2.5\%$  (“var  $< 2.5\%$ ”) and  $< 1\%$  (“var  $< 1\%$ ”), and analyzed these two datasets with the LMEM. We chose two different thresholds for within-block variability. 2.5% was chosen to filter the most varying blocks only and 1% was chosen to ensure negligible variation during the blocks.



Ratios between variability during the blocks for ROIs give insight into the amount of physiological changes present in the StO<sub>2</sub>. Calculation starts with scaled MAD of all samples of a block. Then, the scaled MAD of one ROI is divided by that of the other, and in the next step the mean of the five blocks is calculated. The mean and median over all subjects are shown in Table 3. As this analysis only takes variability within one block into account, i.e., without sensor resiting, these ratios are not affected by this procedure and we would not expect higher variability in FTL StO<sub>2</sub> compared with OL StO<sub>2</sub>. Short-term artifacts have marginal influence on MAD within a block of one minute. We, therefore, associate the numbers in Table 3 with the level of physiological changes contained in the StO<sub>2</sub> of the different ROIs.

### 3 Results

In total, 37 patients were included in this study. Data from two subjects were discarded due to a complete data loss. From the remaining 35 patients, all available data were analyzed. In some of these subjects, it was not possible to obtain sufficiently reliable data and datasets are partially incomplete: only four measurements were performed (ID 19); due to technical problems with the pulse oximeter instrument, SpO<sub>2</sub> recordings are not available (IDs 3 and 18); OL StO<sub>2</sub> data were only available in a single complete measurement and therefore disregarded (ID7).

A typical time series of one patient (ID 12) of the FTL and OL StO<sub>2</sub> (ROI<sub>1+2</sub>) and of SpO<sub>2</sub> is shown in Fig. 2, which shows that all signals are strongly affected by spontaneous fluctuations.

Block medians of SpO<sub>2</sub> that indicate the degree of systemic physiological changes are shown in Fig. 3(a). The SD of SpO<sub>2</sub> throughout the whole recording varies considerably between subjects and is shown in Fig. 3(b). In a number of subjects, there are high values meaning significant variation over time and only in very few cases SD is <1%.

The block medians of StO<sub>2</sub> are graphically presented as box plots in Fig. 4. These show StO<sub>2</sub> for ROI<sub>1</sub>, ROI<sub>2</sub>, and ROI<sub>1+2</sub> of FTL and OL sensors, respectively. In all figures, a number of subjects with a high variability of values were found, whereas for many subjects there is very little variation.

This finding is illustrated even better when plotting the variability of block medians of StO<sub>2</sub>: SD and scaled MAD for the block medians of StO<sub>2</sub> of each subject are shown in Fig. 5. Data are sorted in ascending order and the median values of both are indicated by horizontal lines. SD-sorted  $1.4826 \times \text{MAD}$  allows for subject-wise comparison of SD and scaled MAD.

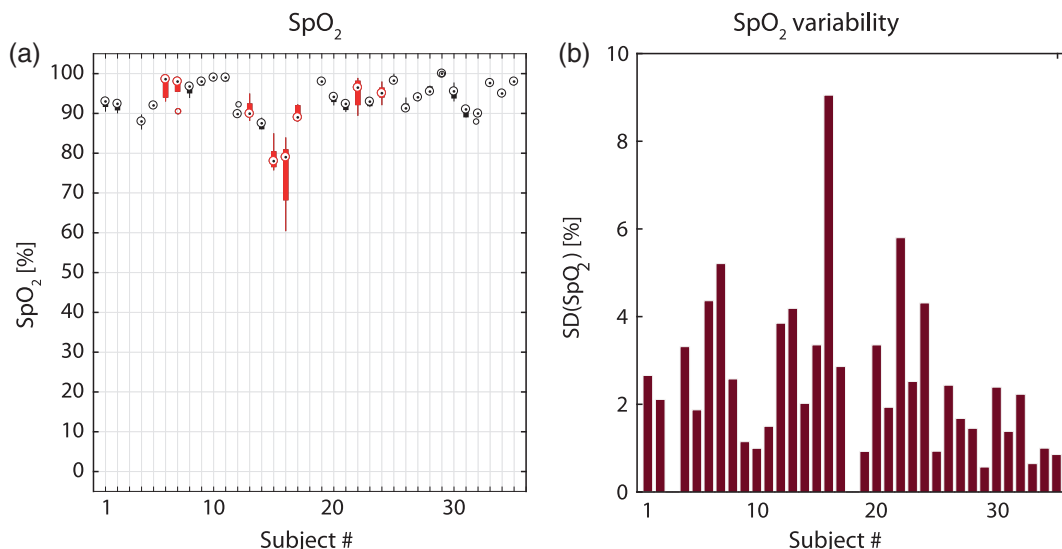
The resulting SD<sub>intraSubj</sub> for StO<sub>2</sub> and SpO<sub>2</sub> obtained by different analyses (I to VI) are shown in Table 2. A number of subjects (subj) and number of blocks considered for each analysis type are provided. Different numbers of subjects with available data or blocks filtered out by the applied criteria are the causes for given spans. For example, “all available” contains FTL data from 35, OL data from 34, and SpO<sub>2</sub> data from 33 subjects, resulting in a span of 33 to 35, respectively.

An analysis of the mean of variability during the 1-min measurements for StO<sub>2</sub> is shown in Table 3.

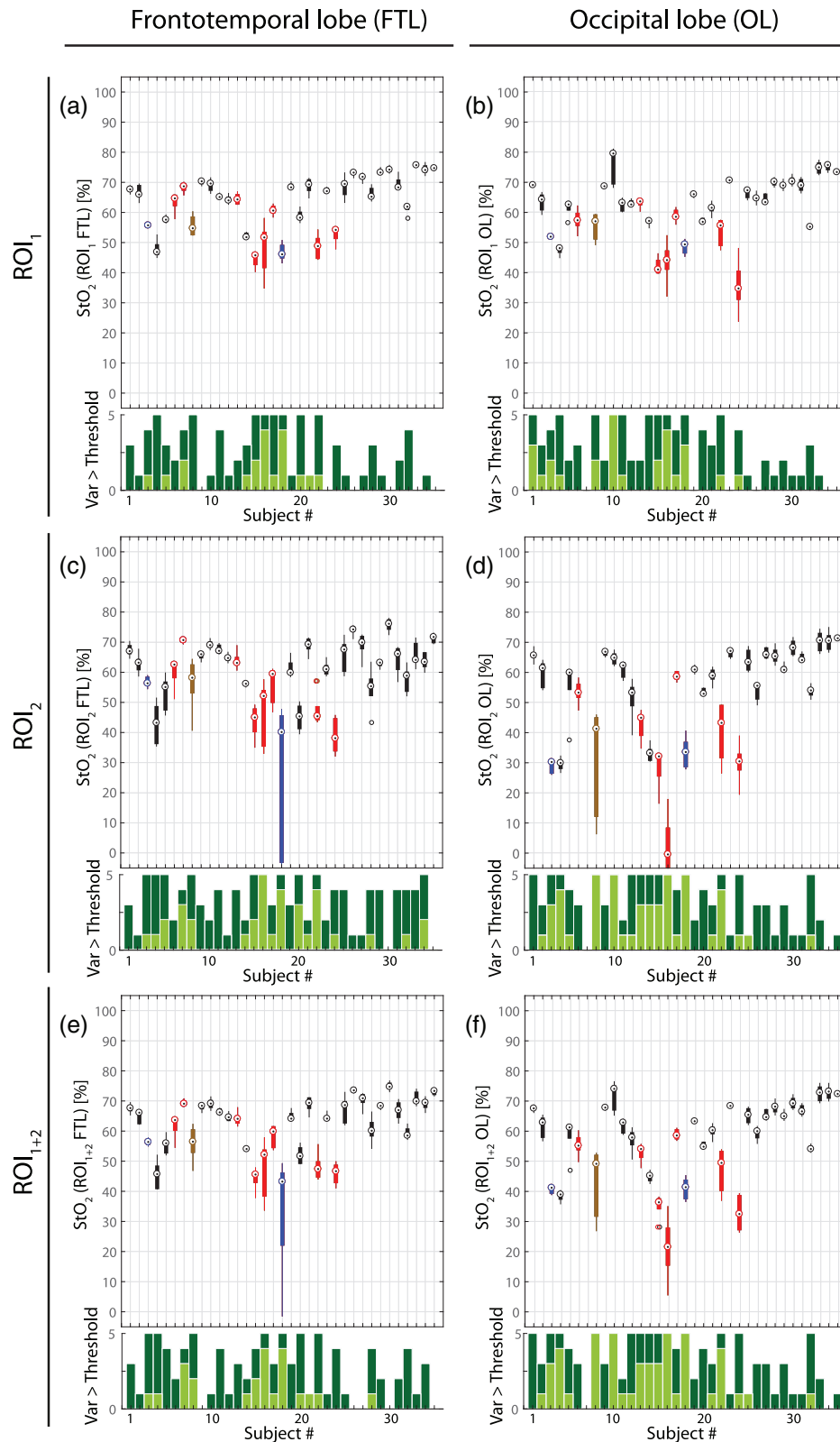
### 4 Discussion

Taking all available data of our precision assessment in clinically stable preterm infants into account results in a variability of 2.64% for OxyPrem. However, we attribute a major stake of this number to physiological changes occurring during the measurement. Different methods for reducing the influence of physiological changes suggest that OxyPrem precision is actually superior and  $\leq 1.85\%$ .

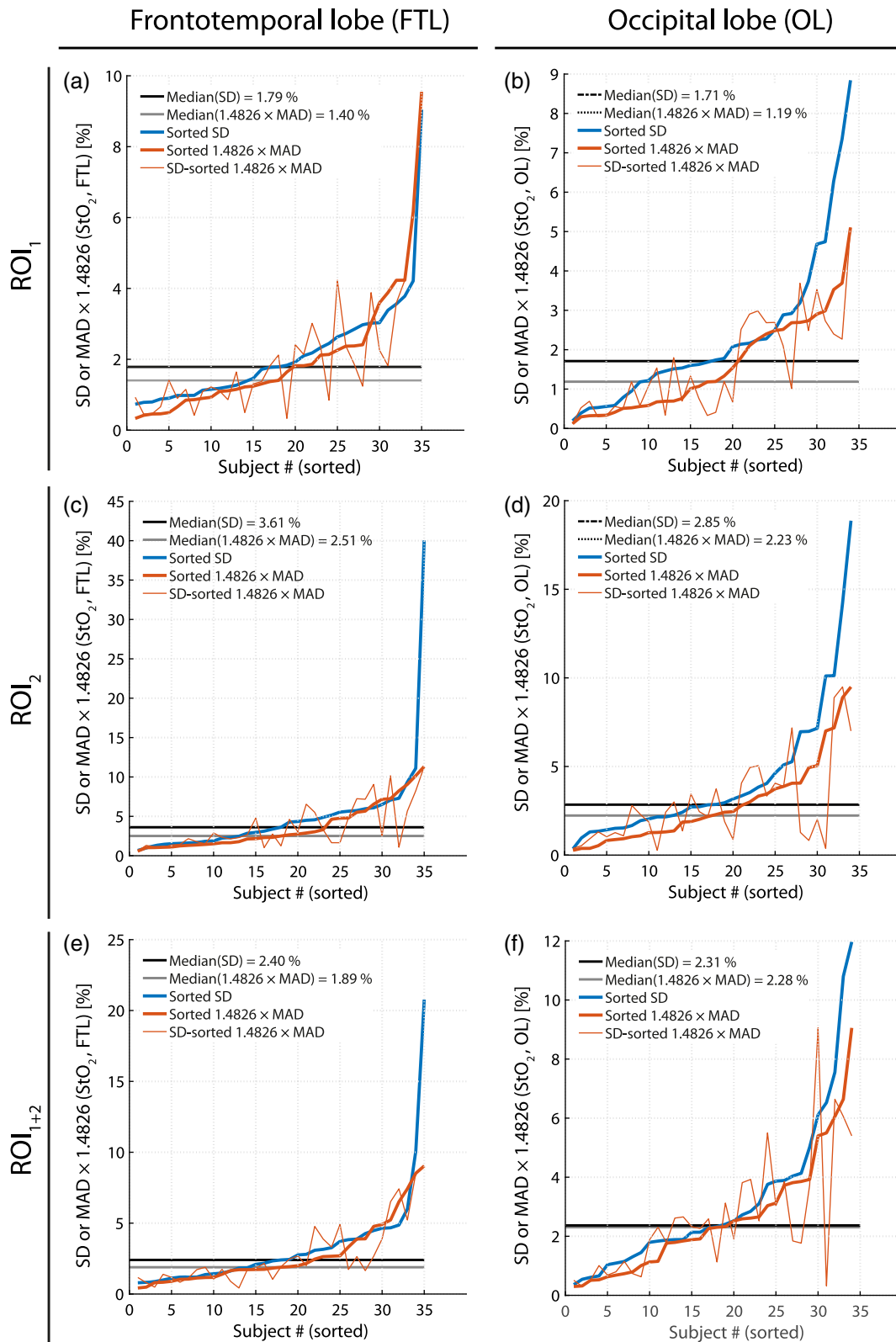
What does precision in cerebral monitoring with NIRS-based oximeters mean for clinical applications? Precision corresponds to the reliability of the numbers displayed to the clinician and is device specific. If an instrument displays an StO<sub>2</sub> of 60% and we assume that the error of measurement is normally distributed around the 60%, we would like to know the probability that the true StO<sub>2</sub> is in reality < 55%, values that are considered to be too low.<sup>19</sup> For an instrument with 5% precision, in 1 out of 6 cases the true StO<sub>2</sub> is indeed < 55%, whereas for 2.5% precision it is only 1 out of 46 and for 2% precision 1 out of 172 cases, respectively. Thus, precision is directly linked to how trustworthy a displayed StO<sub>2</sub> is and it represents a crucial parameter when clinically applying cerebral oximetry.



**Fig. 3** (a) Box plot visualizing the variability between the five block medians of SpO<sub>2</sub>. Subjects with unstable SpO<sub>2</sub> are painted in red. (b) SD of SpO<sub>2</sub> throughout the whole recording for all subjects.



**Fig. 4** Box plots visualizing the variability between the five block medians. Left (a, c, e):  $\text{StO}_2$  of the FTL sensor that was repeatedly removed and reattached (resited). Right (b, d, and f):  $\text{StO}_2$  of the fixed OL sensor. (a and b) Results of  $\text{ROI}_1$ , (c and d) of  $\text{ROI}_2$ , and (e and f) of  $\text{ROI}_{1+2}$ . Manual artifact rating indicated multiple artifacts in five subjects (IDs 15, 16, 17, 18, and 22). Colored data correspond to subjects not having passed the quality criteria for stable physiology due to: missing  $\text{SpO}_2$  data (blue), unstable  $\text{SpO}_2$  (red) or unstable OL  $\text{StO}_2$  (brown). The lower subplots indicate how many out of the typically five blocks per subject exceeded the threshold (T) for variation during these periods (light green: “var > 2.5%”, dark green: “var > 1%”).



**Fig. 5**  $\text{StO}_2$  variability expressed as SD and scaled median absolute deviation (scaled MAD) of the block medians. Data are sorted by ascending SD and scaled MAD. Additionally scaled MAD is displayed with ordering by ascending SD for a subject-wise comparison. Median values of both SD and scaled MAD are indicated by horizontal lines. Left (a, c, and e):  $\text{StO}_2$  of the FTL sensor, which was repeatedly removed and reattached (resited). Right (b, d, and f):  $\text{StO}_2$  of the fixed OL sensor. (a and b) show results of  $\text{ROI}_1$ , (c and d) show results of  $\text{ROI}_2$ , and (e and f) show results of  $\text{ROI}_{1+2}$ .

**Table 2** Variability (Var) between the 1-min measurements (blocks) of SpO<sub>2</sub> and StO<sub>2</sub> of the FTL and OL. Variability expressed as within-subject (SD<sub>intraSubj</sub>) in StO<sub>2</sub> [%] calculated by LMEM based on all data and after applying different criteria (I–IV). For each analysis, the minimal and maximal number of available subjects and blocks passing the quality criteria is given, e.g., II “manual rating” contains FTL data from 30, OL data from 29, and SpO<sub>2</sub> data from 29 subjects, i.e., a span of 29 to 30.

	Method	Subjects (n)	Blocks (n)	ROI <sub>1</sub> FTL [%]	ROI <sub>2</sub> FTL [%]	ROI <sub>1+2</sub> FTL [%]	ROI <sub>1</sub> OL [%]	ROI <sub>2</sub> OL [%]	ROI <sub>1+2</sub> OL [%]	SpO <sub>2</sub> [%]
I	All available	33 to 35	164 to 174	2.64	8.21	4.84	3.06	5.79	4.15	2.35
II	Manual rating	29 to 30	144 to 149	2.02	4.14	2.77	2.76	5.10	3.61	1.44
III	Stable SpO <sub>2</sub>	25	124	1.94	4.18	2.76	2.24	5.13	3.54	1.03
IV	Stable physiology	24	119	1.85	3.82	2.54	2.08	3.53	2.65	1.01
V	Var < 2.5%	30 to 35	120 to 155	2.11	3.93	2.77	2.32	2.86	2.62	-
VI	Var < 1%	20 to 27	54 to 88	2.04	2.65	2.56	1.77	2.70	2.21	-

**Table 3** Ratios of variability during the 1-min measurements.

	ROI <sub>1</sub> OL/FTL	ROI <sub>2</sub> OL/FTL	ROI <sub>1+2</sub> OL/FTL	ROI <sub>1</sub> /ROI <sub>2</sub> FTL	ROI <sub>1</sub> /ROI <sub>2</sub> OL
Median	1.07	1.18	1.20	0.74	0.79
Mean	1.34	1.57	1.42	0.78	0.77

To determine the precision *in vivo*, a sensor is repeatedly placed over the same tissue and the variation of the values between these resittings corresponds to the precision. This is only true, if the measured physiological parameter changes negligibly. Under this assumption, the results show that our new in-house developed NIRS-based oximeter device OxyPrem with a precision of 2.64% achieves the requirements for clinical practice, which are requested to be <3% for cerebral oxygenation of preterm infants.<sup>2,3</sup>

Typically this precision corresponds to 5% for other instruments: e.g., 5.2% for neonates<sup>4</sup> and 6.1% were reported for anesthetized children<sup>20</sup> for the NIRO-300 (Hamamatsu), 6.7% in neonates<sup>21</sup> and 7.1% in anesthetized children<sup>20</sup> for the INVOS 5100C adult SomaSensor (Medtronic), and 4.8% in term born neonates for the INVOS 5100C OxyAlert neonatal sensor.<sup>22</sup> A better repeatability in neonates with SD<sub>intraSubj</sub> of 1.7% was reported<sup>8</sup> for a discontinued device<sup>7</sup> and 2.8% for the FORE-SIGHT neonatal small dual sensor (Casmed).<sup>22</sup> However, for both sensors a decreased dynamic range of changes in StO<sub>2</sub> was reported,<sup>22,23</sup> i.e., a lower sensitivity. SD<sub>intraSubj</sub> of 2.76% in 30 mostly preterm infants was demonstrated for a prototype NIRS device employing a self-calibrating algorithm.<sup>16</sup> By excluding tissue homogeneity, a higher precision of 2.0% in term infants and 4.2% in preterm infants was determined for the OxiplexTS (ISS).<sup>5</sup>

OxyPrem is capable of measuring two independent StO<sub>2</sub> values (ROI<sub>1</sub> and ROI<sub>2</sub>). We expected that the average of these two StO<sub>2</sub> (ROI<sub>1+2</sub>) would perform better than each individual ROI, which is not the case. The reason is that ROI<sub>1</sub> incorporates the longest source–detector distance available (35 mm), which obviously adds stability. We consequently refer to ROI<sub>1</sub> when we speak of the precision of OxyPrem.

Precision is only correctly determined if the true StO<sub>2</sub> does not change between measurements. However, this assumption is

incorrect because the human body constantly regulates and creates physiological changes. So far, these fluctuations were deemed negligible. It is one major result of this study to be the first to quantify these fluctuations and to estimate the respective errors in the precision calculation. Indeed, it was shown that these fluctuations are often not negligible at all. Therefore, an assessment of precision of an instrument requires a correction for this influence.

#### 4.1 Contribution of Spontaneous Physiological Changes

In subject 12, SpO<sub>2</sub> fluctuates with an SD of almost 4% (Fig. 3). Figure 2 shows that in SpO<sub>2</sub> and both StO<sub>2</sub> signals of subject 12, there are significant oscillations present which is due to coupling between SpO<sub>2</sub> and StO<sub>2</sub>.<sup>7</sup> In addition, most of the variation in the FTL StO<sub>2</sub> is not actually caused by the sensor resitting. Furthermore, in the OL StO<sub>2</sub> even stronger changes are visible, although the sensor was kept in place (Fig. 2).

There are different types of contributions to the variability observed in this study. The first type is variations occurring when the true StO<sub>2</sub> is unchanged. This type is defined as the precision of a device, i.e., (1) random noise, (2) changes in shape of the sensor, and (3) differences in optical coupling due to the resitting of the sensor. In an *in vivo* assessment, however, changes in the true StO<sub>2</sub> are present. Some are caused by the sensor resitting: (4) tissue heterogeneity, if the measurement location is not kept exactly the same, (5) changes in skin perfusion and oxygenation caused by the manipulation, and (6) changes in brain perfusion and in arterial and venous compartment sizes because the head was moved during the manipulation. In addition, spontaneous physiological changes may affect cerebral NIRS measurements.<sup>24–27</sup> These occur independently of the sensor’s removal and reattachment process and reflect true alterations in StO<sub>2</sub>: (7) systemic changes in SpO<sub>2</sub> (desaturations and fluctuations and/or oscillations) and (8) changes in blood flow, hemoglobin concentration, and oxygen consumption of the brain due to different regulation processes.

The definition of precision is the random error associated with measuring an unchanged quantity repeatedly. However, most of the contributions to the variability listed above correspond to changes of the true StO<sub>2</sub> (contributions 4 to 8). In the analysis I including all subjects (“all available”) (



(Table 2),  $SD_{intraSubj}$  for ROI<sub>1</sub> OL StO<sub>2</sub>, SpO<sub>2</sub>, and ROI<sub>1</sub> FTL StO<sub>2</sub> are almost the same. This shows that the major contribution to variability cannot be the resiting process and that  $SD_{intraSubj}$  corresponding to the observed variability is systematically higher than the instrument-related variability in repeated measurements (i.e., precision, contributions 1 to 3).

Figure 2 shows another effect that matters for *in vivo* precision experiments. The shorter the measurement periods, the stronger is the effect of physiological fluctuations. In this subject (ID 12), one period of physiological fluctuation is ~15 s. Due to 1-min averaging, these changes cancel to a certain degree. Therefore, although SD of SpO<sub>2</sub> is ~4% (Fig. 3), this measurement passed the criteria for stable physiology (“stable physiology”) with an SD of 1.1% between the block medians. With shorter block sizes, this smoothing effect diminishes, resulting in higher  $SD_{intraSubj}$ . This may explain the different precisions determined by different research groups for the same sensor on the same patient groups.

In Table 3, the ratio for OL/FTL is >1 in all three cases (ROI<sub>1</sub>, ROI<sub>2</sub>, and ROI<sub>1+2</sub>) and for ROI<sub>1</sub>/ROI<sub>2</sub> is <1 in both cases (FTL and OL). As MAD is robust to short-time artifacts (e.g., motion artifacts), we assume that the variation during the blocks originates from spontaneous physiological fluctuations in the StO<sub>2</sub> (Fig. 2). Generally, physiological variation in StO<sub>2</sub> in the OL is higher than in the FTL. In addition, we see less variation in StO<sub>2</sub> of ROI<sub>1</sub> than in ROI<sub>2</sub>. ROI<sub>1</sub> with SDS of 15 and 35 mm collects data from deeper layers of the brain and averages a larger volume of tissue, leading to a more stable signal.

#### 4.2 How to Reduce the Effect of Spontaneous Physiological Changes in Case of NIRS Reproducibility Measurements

For correctly calculating precision, the true StO<sub>2</sub> needs to be unchanged. However, several subjects showed strong physiological changes during the measurement, invalidating this assumption. This causes a systematic overestimation of the precision. To minimize this effect, we applied different criteria to remove individual blocks or subjects and analyzed the remaining data with the LMEM. Table 2 shows the resulting  $SD_{intraSubj}$  and the number of subjects and blocks, which have passed the respective criteria. For clinical application, there is of course no requirement for StO<sub>2</sub> to remain unchanged and a patient with variable physiology can be reliably monitored with a NIRS instrument. Such patients probably are among those benefiting the most from such monitoring.

- II. The approach with manual artifact rating depends on the person performing the rating. This is a difficult task because not only movement artifacts but also continuous spontaneous physiological changes have to be considered. We, therefore, do not favor this approach.
- III. The approach to include subjects with stable systemic physiology (“Stable SpO<sub>2</sub>”) requires additional sensors in place (SpO<sub>2</sub>). The criterion allows periodic fluctuations within the blocks but requires the physiology to be stable throughout the whole experiment. This approach results in smaller  $SD_{intraSubj}$  (Table 2). In at least one subject (ID 8), a substantial slow trend in StO<sub>2</sub> was observed, which clearly violates the assumption of stable “true” StO<sub>2</sub>. This infant was

not identified by the SpO<sub>2</sub> criterion. This shows that although both are coupled,<sup>7</sup> SpO<sub>2</sub> cannot predict cerebral StO<sub>2</sub>. Therefore, monitoring systemic physiology alone is not sufficient to guarantee stable cerebral StO<sub>2</sub>.

- IV. Thus, we added another criterion based on variability of OL StO<sub>2</sub>. We consider the remaining subjects having “stable physiology,” which is reflected in even lower  $SD_{intraSubj}$  (Table 2), than in the previous approaches. However, precision is probably still considerably overestimated due to a major physiological contribution because  $SD_{intraSubj}$  of StO<sub>2</sub> of the FTL sensor is almost identical to  $SD_{intraSubj}$  of the OL sensor which was not resited.
- V. and VI. The approach limiting variability within each block (“var < 2.5%” and “var < 1%”) discards all blocks with periodic fluctuation irrespective whether or not they influence the block median. However, a slow change in oxygenation is not removed and still causes an overestimation of  $SD_{intraSubj}$ . Especially the strict criterion with variability <1% discards many blocks, with only approximately three left per subject on average, which reduces statistical power.

The box plot in Fig. 4 and the sorted variabilities of all subjects in Fig. 5 show that the majority of subjects have relatively small deviation between the blocks and only few subjects considerably increase the overall  $SD_{intraSubj}$ . Generally, SD and scaled MAD are similar for normally distributed data. Our dataset is not normally distributed because scaled MAD provides on average lower values than SD. The reason is that MAD provides better robustness to outliers, whereas  $SD_{intraSubj}$  is only valid for normally distributed data. This is further demonstrated by median values of scaled MAD in Fig. 5 being typically lower than the  $SD_{intraSubj}$  of LMEM (Table 2). Therefore, even in datasets with strong physiological variability excluded, LMEM systematically overestimates the variability. This also means that the precision of OxyPrem is better than 1.85% in reality. Although the resulting variability still depends on physiological changes and is still higher than the true precision of the instrument, it definitely is closer to the true precision.

For all discussed methods to obtain the true precision,  $SD_{intraSubj}$  of the repositioned sensor was substantially lower than in the analysis based on all available data. This indicates that physiological changes contributed a major part to  $SD_{intraSubj}$ . Neglecting these physiological changes leads to a substantial overestimation of the precision. The proposed methods lead to a much reduced and more correct precision, which is still somewhat overestimated. For future precision studies, we recommend employing the method requiring stable SpO<sub>2</sub> with some additional criteria to detect trends in StO<sub>2</sub>, as both systemic and cerebral physiology are required to be monitored and restricted to certain limits, to reduce their influence on the precision determination. We thus refer to ROI<sub>1</sub> results with approach “stable physiology” (IV) as best estimate for OxyPrem precision.

#### 4.3 Strengths and Limitations of this Study

A strength of this study is that we clearly demonstrated how systemic physiological changes have an impact on precision

assessments of cerebral oximeters in clinically stable preterm infants. We further presented several approaches to cope with this undesired effect; all of them estimate precision more realistically than approaches neglecting the effect of physiological changes. A limitation of this study is that we acquired data only in preterm infants. We, therefore, cannot conclude a dominant influence of physiological changes in other patient groups, although it is very likely that this factor plays also a significant part for reproducibility studies based on other populations. We measured SpO<sub>2</sub> with a clinical pulse oximeter for the approaches applying thresholds on SpO<sub>2</sub> variability (III and IV). We consider this pulse oximeter to be reliable in particular as we only considered changes in SpO<sub>2</sub> and since the sensor was not resited. Therefore, random errors due to sensor placement or a device-specific SpO<sub>2</sub> bias do not affect the exclusion of subjects.

Do FTL and OL StO<sub>2</sub> represent cerebral signals? Influence of extracerebral tissues has been reported for cerebral measurements in adults<sup>28,29</sup> and could potentially also affect our study. However, the neonatal skull and other extracerebral layers are thin<sup>30</sup> and are therefore less relevant. Furthermore, OxyPrem employs a multidistance algorithm that effectively cancels influence of superficial tissue. Therefore, we can safely state that OxyPrem StO<sub>2</sub> reflects cerebral oxygenation in neonates. OxyPrem has similar deep-tissue sensitivity as common adult oximeters.<sup>12</sup>

## 5 Conclusions

We presented a methodology to assess precision of NIRS-based oximetry devices. For the first time, we showed that indeed in neonates, systemic physiology and cerebral regulation cause major changes in StO<sub>2</sub> that invalidate conventional precision estimates. Therefore, in addition to StO<sub>2</sub> of the sensor which was removed and reattached, SpO<sub>2</sub> and StO<sub>2</sub> need to be measured continuously to identify such variations. It is necessary to correct for this type of variability. Several such methods were presented. For future studies, we recommend using an analysis that limits influence of systemic and cerebral physiology to obtain a proper precision estimate. OxyPrem has shown to be a highly precise instrument with a precision  $\leq 1.85\%$ , which is better than instrumentation currently used in clinics.

## Disclosures

Dr. Kleiser, Mr. Ostojic, Dr. Nasserli, and Prof. Wolf are inventors of a pending European patent filed by University of Zurich, which covers the features related to the superior precision of OxyPrem.

## Acknowledgments

This work was funded by the Nano-Tera projects ObeSense, and NewbornCare, the Swiss National Science Foundation (project 159490), and the BioEntrepreneur-Fellowship of the University of Zurich, Reference No. BIOEF-17-004.

## References

1. J. J. Volpe, "Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances," *Lancet Neurol.* **8**(1), 110–124 (2009).
2. G. Greisen et al., "Cerebral oximetry in preterm infants: an agenda for research with a clear clinical goal," *Neurophotonics* **3**, 031407 (2016).
3. M. Wolf and G. Greisen, "Advances in near-infrared spectroscopy to study the brain of the preterm and term neonate," *Clin. Perinatol.* **36**(4), 807–834, vi (2009).
4. L. C. Sorensen and G. Greisen, "Precision of measurement of cerebral tissue oxygenation index using near-infrared spectroscopy in preterm neonates," *J. Biomed. Opt.* **11**(5), 054005 (2006).
5. S. J. Arri et al., "Precision of cerebral oxygenation and hemoglobin concentration measurements in neonates measured by near-infrared spectroscopy," *J. Biomed. Opt.* **16**(4), 047005 (2011).
6. M. Wolf et al., "Tissue oxygen saturation measured by near infrared spectrophotometry correlates with arterial oxygen saturation during induced oxygenation changes in neonates," *Physiol. Meas.* **21**(4), 481–491 (2000).
7. S. Hyttel-Sorensen et al., "Tissue oximetry: a comparison of mean values of regional tissue saturation, reproducibility and dynamic range of four NIRS-instruments on the human forearm," *Biomed. Opt. Express* **2**(11), 3047–3057 (2011).
8. J. Menke et al., "Reproducibility of cerebral near infrared spectroscopy in neonates," *Biol. Neonate* **83**, 6–11 (2003).
9. D. M. Hueber et al., "New optical probe designs for absolute (self-calibrating) NIR tissue hemoglobin measurements," *Proc. SPIE* **3597**, 618–631 (1999).
10. F. Scholkmann, A. J. Metz, and M. Wolf, "Measuring tissue hemodynamics and oxygenation by continuous-wave functional near-infrared spectroscopy—how robust are the different calculation methods against movement artifacts?" *Physiol. Meas.* **35**(4), 717–734 (2014).
11. S. Kleiser et al., "Comparison of tissue oximeters on a liquid phantom with adjustable optical properties," *Biomed. Opt. Express* **7**(8), 2973–2992 (2016).
12. N. Nasserli et al., "Quantifying the effect of adipose tissue in muscle oximetry by near infrared spectroscopy," *Biomed. Opt. Express* **7**(11), 4605–4619 (2016).
13. J. Pinheiro et al., "NLME: linear and nonlinear mixed effects models," R package version 3, pp. 1–137, <https://CRAN.R-project.org/package=nlme> (2018).
14. N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics* **38**(4), 963–974 (1982).
15. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/> (2018).
16. C. Jenny et al., "Reproducibility of cerebral tissue oxygen saturation measurements by near-infrared spectroscopy in newborn infants," *J. Biomed. Opt.* **16**(9), 097004 (2011).
17. F. R. Hampel, "The influence curve and its role in robust estimation," *J. Am. Stat. Assoc.* **69**(346), 383–393 (1974).
18. P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. Am. Stat. Assoc.* **88**(424), 1273–1283 (1993).
19. S. Hyttel-Sorensen et al., "Cerebral near infrared spectroscopy oximetry in extremely preterm infants: phase II randomised clinical trial," *Br. Med. J.* **350**, g7635 (2015).
20. A. Dullenkopf et al., "Measurement of cerebral oxygenation state in anaesthetized children using the INVOS 5100 cerebral oximeter," *Paediatr. Anaesth.* **13**, 384–391 (2003).
21. S. Hyttel-Sorensen et al., "A comparison between two NIRS oximeters (INVOS, OxyPrem) using measurement on the arm of adults and head of infants after caesarean section," *Biomed. Opt. Express* **5**(10), 3671–3683 (2014).
22. T. W. Hessel, S. Hyttel-Sorensen, and G. Greisen, "Cerebral oxygenation after birth - a comparison of INVOS and FORE-SIGHT near-infrared spectroscopy oximeters," *Acta Paediatr.* **103**, 488–493 (2014).
23. M. Wolf et al., "Testing near infrared spectrophotometry using a liquid neonatal head phantom," *Proc. SPIE* **3566**, 79–86 (1998).
24. F. Scholkmann et al., "End-tidal CO<sub>2</sub>: an important parameter for a correct interpretation in functional brain studies using speech tasks," *Neuroimage* **66**, 71–79 (2013).
25. I. Tachtsidis and F. Scholkmann, "False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward," *Neurophotonics* **3**(3), 031405 (2016).
26. F. Scholkmann, M. Wolf, and U. Wolf, "The effect of inner speech on arterial CO<sub>2</sub> and cerebral hemodynamics and oxygenation: a functional NIRS study," *Adv. Exp. Med. Biol.* **789**, 81–87 (2013).

27. M. Caldwell et al., "Modelling confounding effects from extracerebral contamination and systemic factors on functional near-infrared spectroscopy," *Neuroimage* **143**, 91–105 (2016).
28. T. J. Germon et al., "Near-infrared spectroscopy in adults: effects of extracranial ischaemia and intracranial hypoxia on estimation of cerebral oxygenation," *Br. J. Anaesth.* **73**(4), 503–506 (1994).
29. T. J. Germon et al., "Extracerebral absorption of near infrared light influences the detection of increased cerebral oxygenation monitored by near infrared spectroscopy," *J. Neurol. Neurosurg. Psychiatry* **58**(4), 477–479 (1995).
30. A. Demel et al., "Correlation between skin, bone, and cerebrospinal fluid layer thickness and optical coefficients measured by multidistance frequency-domain near-infrared spectroscopy in term and preterm infants," *J. Biomed. Opt.* **19**(1), 017004 (2014).

Biographies for the authors are not available.